HAS QUANTITATIVE RISK ASSESSMENT BEEN OF BENEFIT TO THE SCIENCE OF TOXICOLOGY?

2006 American College of Toxicology Distinguished Service Award Lecture

Dr. Joseph V. Rodricks, D.A.B.T.
ENVIRON International Corporation, Arlington VA

Quantitative approaches to evaluating the risks of chemical toxicity entered the lives of toxicologists in the mid- 1970's, and the continuing interaction of toxicology and risk assessment has been of benefit to both disciplines. I will summarize the origins of the interaction, the reasons for it, and the difficult course it has followed. In doing so, I will set the stage for a discussion of how the type of thinking that informs risk-based decision-making provides important benefits to the continuing development of the science of toxicology. There will continue to be societal pressure for the development of reliable knowledge about the public health importance of the enormous variety of chemical exposures we all incur, from conception to death. Risk assessment is the framework used to organize and convey that knowledge. Toxicology is the principle discipline used to give scientific substance to that framework. Social acceptance of every manifestation of the modern chemical age requires high assurance that the public health is not threatened, and that assurance depends upon continued improvements in these two mutually dependant disciplines.

On a late Friday afternoon in the summer of 1974, I spent an hour with Dr. Leo Friedman in his office at the Food and Drug Administration (FDA). Leo was then director of the agency's toxicology division. He had asked me to visit with him to discuss a report I'd written expressing dismay over what I perceived to be deficiencies in the agency's approach to regulating carcinogens in general, and aflatoxins in particular. Although I did not realize it at the time, my hour with Leo Friedman—someone who had begun his FDA career as an animal caretaker and who had worked his way through college and graduate school while holding on to his FDA day job—set me onto the professional path I have followed for the past 30 years. That path has involved the pursuit of risk-based decision-making to protect the health of humans from the toxic effects all substances can express when exposures become excessive.

This singularly gratifying award from the ACT, to which I will return at the end of my lecture, provides me a most welcome opportunity to outline the history of that pursuit, what it has meant for the science all of us at this meeting are engaged in, and for the public health and regulatory policies that surround it. I remain a strong advocate for risk-based decision-making, but as I shall try to make clear in this paper, I am not blind to the problems it continues to encounter, and that we need to strive to overcome.

It is no doubt a biased history I relate, but this might be excusable because I am not a professional historian. As my title suggests, I am also going to risk saying something about whether the science we are all most interested in advancing has benefited from the kind of thinking that drives risk assessment.

At the time of my 1974 discussion with Leo Friedman, I had been a laboratory scientist at the FDA for 9 years, and during that time had pursued and received a doctorate in biochemistry at the nearby University of Maryland. Except for the year of post-doctoral research I did at Berkeley, where I worked on the problem of paralytic shellfish poison, aflatoxins and some other mycotoxins had been my subjects of study. I came to the agency soon after the discovery of the aflatoxins and was quickly involved in the isolation and purification of these compounds, and in the production of radiolabeled aflatoxins for purposes of tracking their metabolic pathways. By the end of my first year at the agency, the hepato-carcinogenicity of aflatoxins had been established in six different animal species, and it wasn't too long before some suggestive evidence regarding human carcinogenicity emerged from studies in Thailand and South Africa (Rodricks et al., 1977). At the time no known chemical had produced malignancies in animals at doses as low as those at which aflatoxins (B1 in particular) were active; even today it is surpassed in this respect only by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin. (A most informative exercise involves comparing and contrasting this pair of extraordinarily potent carcinogens, which otherwise differ in so many ways).

Aflatoxins were, during this same period, found to be not uncommon contaminants of certain human foods. The molds that produced them in raw agricultural products could be destroyed, or at least disguised, during processing, but aflatoxins proved sufficiently stable to survive processing. Convincing evidence of carcinogenicity coupled with certain knowledge that people were being exposed, on a worldwide and regular basis at doses that were not obviously trivial, placed these naturally occurring compounds squarely under the public health spotlight. We estimated, for example, that typical human exposures in the United States were about 500-1,000 times less than the minimum carcinogenic dose level, and that "high end" exposures were only about 100 times less (Stoloff and Rodricks, 1977).

At that time, FDA toxicologists and most others in the federal government believed that no safe levels of exposure could be identified for carcinogens, or (perhaps more significantly) that no safe levels of exposure actually existed at any dose greater than zero. These views rested upon a long series of expert government reports from the National Cancer Institute (IRLG, 1979). Congressional testimony from these experts strongly influenced legislative thinking during the 1950's, and led, among other things, to the famous Delaney Amendment, a feature of the federal food law that forbids the deliberate introduction into food of any carcinogen, human or animal (NRC, 1983).

Aflatoxin was not a deliberately introduced food ingredient, but a product of mold growth (a mycotoxin), so the FDA regulators argued that the agency had no authority under the law to ban aflatoxin. This was perhaps fortunate because aflatoxin could not be banned unless the very foods of which it was a routine contaminant were themselves banned, or at least disposed of in large quantities. But given the absence of an identifiable "safe" level, it seemed that something had to be done to limit human exposures to aflatoxins.

The toxicologists having given them only bad news and no way to deal with it, the regulators in the FDA turned to the analytical chemists. Thus came regulation by the odd rule that says: "If I can detect it, it is dangerous; but if I cannot detect it, it's not harmful." The chemists could demonstrate, in the 1965-1967 period, that aflatoxins could be measured and confirmed to be present in foods at 30 μg/kg and above, but below that level aflatoxin, even if present, could not be measured and confirmed to be present. A regulatory action level could thus be established at 30 μg/kg (Rodricks, 2006). No one, of course, actually believes the "odd rule", but regulation based on analytical detection limits has the same effect as would believing it!

Food surveys under that action level turned up contamination in only a small percent of samples examined. Those samples could be removed from commerce without significant economic damage. But the analytical chemists kept improving their methods—they are always aiming at the unachievable "zero"—and it wasn't long before the FDA had to reduce the action level to 20 µg/kg. It was, of course, completely predictable that as analytical detection limits fell, aflatoxin would be found in greater and greater fractions of affected foods. When the chemists declared in 1969 that 2 µg/kg was easily achievable, the FDA knew that regulation based on such a detection limit would have a devastating effect on affected segments of the food supply. A model for regulatory decision-making that relied on what the chemists could reliably measure, or on some other factor unrelated to the toxic properties of the substance being regulated, was neither rational nor useful. For some substances a regulatory standard that was based on analytical detection limits might be quite protective of health, but in other cases it might be seriously deficient in this respect. Because the chemists' abilities to measure a chemical is completely unrelated to any risks to health that substance might pose, regulatory standards based on that measurement capability are of completely unknown health protective value. The very same argument holds for regulatory policies that seek to protect people by developing standards based solely on criteria such as "maximum achievable control technologies." There is no relationship between these achievable control levels (however they might be defined) and the likelihood of adverse health effects posed by chemicals controlled at those levels. Of course, as the chemists and the control engineers continue to chase after "zero," standards have to be continually reduced, again without an understanding of whether or not they have been reduced sufficiently to protect the health of exposed humans, or whether they have become unnecessarily restrictive (and therefore unnecessarily burdensome in an economic sense). This problem is not restricted to aflatoxins or other food contaminants, but to most contaminants of the environment (NRC, 1994).

My brief report to Friedman laid out this argument, and included a rather arrogant attack on the failure of the toxicology community to face up to this dilemma. Government and other testing programs had begun to turn up new carcinogens every month, including many commercial products of substantial importance. In a limited number of cases in which the Delaney Amendment could be applied, banning was the easy course. But the government was going to have to deal in some systematic manner with increasing numbers of environmental contaminants and commercial products that were identified as animal carcinogens, and a smaller but significant number that were identified as carcinogens through epidemiology studies.

Leo Friedman told me my three-page memo echoed some ideas he had been thinking about for several years. He told me that the method for establishing safe levels for most chemicals, devised by his predecessor at the FDA, Arnold Lehman, and another FDA toxicologist, O. Garth Fitzhugh, was based on a widely accepted view, held by most toxicologists, that the toxic properties of most chemicals expressed themselves only after a threshold dose was exceeded, and that concept provided a scientific basis for establishing safe levels for humans. The Lehman-Fitzhugh approach, published in the 1949-1955 period, relied upon the application of what were then called "safety factors" to data obtained from experimental toxicology studies and, in some cases, epidemiology studies (Fitzhugh, 1949; Lehman et al 1955). But Friedman was aware that there was a community of experts working in the areas of chemical and radiation carcinogenesis who had developed quite different views of the biological actions of agents having carcinogenic properties, and that their notions of thresholds, reversibility, and dose-response were radically different from those of the traditional toxicologists (Armitage and Doll, 1954).

Leo Friedman pulled from his file a thin folder containing half a dozen publications he asked me to study. He felt it was time for the FDA to find a more scientifically satisfactory way to deal with carcinogens. I promised to return after I had gone through those few papers.

I did indeed study the papers Leo Friedman had given me, but I never returned to discuss them with him. Leo died of heart failure on the weekend after our meeting, at the age of 52.

I can't begin to describe the uninterrupted flow of creative thinking about our science that emerged from the man, and I have always thought he was a much greater loss to our community than we have recognized. But I saw, in reading the several papers he had given me, what he had been thinking about as a way to grapple with the "no safe level" thinking that had become attached to carcinogenic substances. Why not, two of these papers suggested, simply assume that carcinogens were indeed risky "all the way down?" Why not adopt the thinking that motivated the famous and influential 1954 publication of Sir Peter Armitage and Sir Robert Doll that laid out the multistage model of carcinogenesis (Armitage and Doll, 1954)? Here we begin to see carcinogenesis as a probabilistic process, one in which each exposure to some minimally biologically active dose of a carcinogen is sufficient to increase the probability—or risk—that the multi-step carcinogenic process goes to completion. What if we accept these ideas and try to model the dose-risk relationship "all the way down," with the assumption that the risk of carcinogenesis will go to zero only at zero dose?

Risk increases as dose increases, but we should be able to recognize that there are ranges of doses greater than zero at which risks are very small, well within the range of what most rational people would consider so small as to be equivalent to the common-sense notion of safety. This thinking dovetailed neatly with some long-standing ideas about just what we mean when we use the term "safe." The dictionary defines "safe" as "free from harm." But do we know of any activity that we can claim to be completely free from harm, completely without risk? Lots of activities, indeed lots of exposures to chemical products, appear to be free from risk, but none is demonstrably so. We cannot demonstrate the existence of a completely negative condition, a true zero, at least by any known scientific method.

The prominent biostatistician, Nathan Mantel, and an associate, William Bryan, both at the National Cancer Institute, authored one of the papers in Leo's collection. The exposition was clear and the ideas stimulating. The paper included a demonstration of how carcinogenicity dose-response data on a series of polycyclic aromatic hydrocarbons could be described with a simple probit model, and how the tail of the model could be extended well below the typical low end of such curves (which empirically describe risks of tumor development no less than about one-in-10), to estimate doses corresponding to some extremely low and completely unmeasurable risks. Mantel's risk target was one-in-100 million, and the calculated doses corresponding to that excess lifetime risk he labeled "virtually safe." He accepted the absolute meaning of the term "safe," but the adjective he attached to it suggested he considered lifetime risks of one-in-100 million, even if the risk was one of cancer, to be indistinguishable from what most of us would consider to be a state of safety. I should add that Mantel, in extending the trail of the probit model, imposed an artificial slope on it that he thought would place an "upper bound" on the low dose risk, so that risk would not be underestimated; it was intended to impose a "conservative" element into the procedure for extrapolation into the unknown (Mantel and Bryan, 1961).

I was not capable of giving the Mantel-Bryan paper a thoroughly critical reading, but it seemed to me that the approach described gave us a chance to gain some appreciation of the health risks carcinogens might pose at low (human) doses, and to set standards based on the notion that once risks reached some very low levels, we could declare that exposures in these ranges and below it were likely not to threaten health. Decision-making would not be deterministic ("safe/not safe") but rather probabilistic (*i.e.*, risk-based); moreover, it would become possible to gauge the magnitude of risk reductions achieved as regulatory standards were tightened, so that the policy-makers could examine the important question of whether the public health benefits achieved (*i.e.*, risk reduction) by the imposition of various control technologies were worth the costs of achieving them. It appeared that a systematic means for dealing with the increasing numbers of carcinogens that could be found in the environment was available, and that it should be developed for practical application. Decision-making would be linked to the risk characteristics specific to individual carcinogens, together with other factors that dictate the practical limits of risk reduction technologies (Rodricks, 1981).

I spent several months in late 1974 scrutinizing the many animal studies then available on the carcinogenicity of aflatoxins, and exploring the implications of applying the so-called Mantel-Bryan procedure for low-dose extrapolation; I applied other statistical procedures as well. I tried to estimate how human intakes of afla-toxins would decline as regulatory tolerance levels were made more restrictive; having those intake estimates (the human dose) allowed estimation of the risk reductions that might be achieved as tolerances declined. I worked with FDA policy officials to begin to craft a risk-based tolerance for aflatoxins in peanut products.

Because of my work with aflatoxins, I was asked to join another agency effort to move toward risk-based deci-sions. This effort concerned not food contaminants such as aflatoxins, but rather a class of intentionally added substances—drugs used in animals used as food. The uses of those veterinary drugs would result in their presence as "residues" in meat, milk, or eggs. Because they were intentionally added substances, the original form of the Delaney Amendment applied to any such drug that turned out to be carcinogenic. During the 1960's our Congress enacted legislation to allow the use in food production of animal drugs that happened to be carcinogenic. That legislation permitted such use under the condition that "no residue" (the language of the law) of the drug was detectable in human food. This modification of the law came to be known as the "DES Proviso," because issues surrounding the widely used animal growth promoter—and well-established carcinogen—diethylstilbestrol had been the immediate occasion for this legislative change (Hutt and Merrill, 1991).

Here we see immediately the dilemma created by the "no residue" provision: The meaning of that phrase depends upon the limit of detection (LOD) of the analytical method used to search for residues, and that LOD has no bearing at all on the health risk undetected residues (which could be present at any level up to the LOD) might pose. It was conceivable that residues of a highly potent carcinogen might be permitted at higher concentrations in food than those of a far less potent carcinogen, simply because the LOD for the former turned out to be at a much higher concentration that the LOD for the latter. (We can assume residues to be present in concentrations ranging up to the LOD, when "no residue" is found).

It turns out that at this time the FDA was blessed with a very astute general counsel, Peter Barton Hutt, and he had also spent some time with Leo Friedman. Peter had come to lead the effort to put risk assessment into the equation for this class of added food ingredients. Hutt proposed that "safe doses" for carcinogens such as DES could be defined as those associated with lifetime risk levels of less than one-in-one million, when those risks were estimated using a linear, no-threshold model (several publications had demonstrated that the Mantel-Bryan approach could not be counted on to place an upper bound on risk at low doses but that a linear, no-threshold model could). Analytical methods for animal drugs developed by those seeking FDA approvals would be acceptable only if it could be shown, on a drug-specific basis, that their use confirmed the absence of residues at a level that would yield a human dose no greater that the "safe" one. This "Sensitivity of the Method" regulation became the first formally to adopt a risk-based approach for carcino-gens (FDA, 1977). The FDA, after a protracted administrative hearing, acted to extinguish the veterinary uses of DES in 1979, based in part on the fact that the levels of residues detected in food did not meet these new risk-based safety criteria (Zervos and Rodricks, 1982; Rodricks, 1990).

Although it was estimated in a different way than that proposed by Mantel and Bryan, their "virtually safe dose" became Peter Hutt's "safe dose" (FDA, he used to say, did not permit doses for added carcinogens that were only "virtually" safe). The selection of a lifetime risk level considered sufficiently low to define a safe dose was not a scientific, but rather a policy decision.

The EPA wrote guidelines for the conduct of carcinogen risk assessment and published them in 1977 (Albert et al., 1977). The agency had many more carcinogens to deal with than did the FDA, and moved quickly under the leadership of Elizabeth Anderson and the agency's superb consultant, Roy Albert, then at NYU, to develop the technical resources to implement the agency's guidelines in its many programs. Kenny Crump published his so-called "linearized" multistage model for low-dose extrapolation in 1976, and the EPA officially adopted it (IRLG, 1979; EPA, 1986). In doing so, the agency had to find a way to deal with the proliferation of statistically and biologically based models proposed for low-dose extrapolation that was seen in the scientific literature during the 1975-1980 period (NRC, 1983). None could be demonstrated to yield an accurate estimate of low-dose risk, and it could be readily seen that the models under discussion could yield very large differences in low-dose risks for the same carcinogen at the same dose. There was no way to resolve this problem without making a choice among models that was not based on purely scientific under-standing. The linearized multistage model was selected because it seemed to have some basis in the leading mechanistic hypotheses regarding the carcinogenic process, and also because it seemed highly likely that the model—because of its "linearization" at low dose—would not underestimate low-dose risk, that it would, in fact, place an upper bound on low-dose risk. Actual risk might be as large as the upper bound, but could be lower and could even be zero. It is not the case that risk assessors, at least those who truly understood the problem of low-dose extrapolation, have ever claimed that risks predicted in this fashion are known to be accurate, even ignoring the uncertainties introduced by the fact that most risk assessments are based on animal, not human, data. The very small risks estimated by these methods are, in fact, generally unverifiable by any existing epidemiological method (IRLG, 1979, Rodricks, 2006).

These EPA and FDA efforts were intended to make more rigorous and systematic the approach to evaluating carcinogens and to setting standards that took into account the biological properties of the always growing list of contaminants and commercial substances identified as carcinogenic. At the same time quantitative risk assess-ment tools were being developed and used by EPA and FDA, OSHA initiated a massive program to regulate workplace carcinogens. Interestingly, this effort, advanced by Eula Bingham, at first rejected the use of quantita-tive risk assessment (Merrill, 2003). The OSHA interpreted its legal mandate as calling for standard setting in the workplace that was based on the best available technological controls that were economically feasible. Once a substance was identified as a carcinogen, factors other than the risk it posed dictated the Permissible Exposure Limit (PEL). In its effort to regulate benzene under the approach, the agency was challenged by the American Petroleum Institute, and the case rose to the US Supreme Court. The Court directed OSHA to incorporate quanti-tative risk assessments into its regulatory efforts on carcinogens, and required the agency to demonstrate that a significant risk of carcinogenicity existed at the current PEL, and that a significant reduction in risk would occur if a new and lower PEL were instituted. How could a new PEL be justified, the Court reasoned, unless it provided significant health benefits—*i.e.*, risk reductions. The Court explicitly recognized the significant uncertainties asso-ciated with the risk assessment process, but held that OSHA had to do the best it could (Merrill, 2003). OSHA has followed this approach ever since (although most of its efforts on workplace carcinogens took place before the 1990's, and the agency has not been particularly active in this area since that time).

These various regulatory efforts of the late 1970's were met with much skepticism. Some declared that the "no safe level" principle for carcinogens was inviolate, and that risk-based approaches threatened that principle and weakened public health protection (NRC, 1983). Those who attacked from this perspective seemed unconcerned that alternative approaches to dealing with carcinogens (other than banning, which could have only very limited applicability) were far less satisfactory. From the other side came scientific attacks on the "no-threshold" concept, and on the perceived failure of regulatory risk assessors to incorporate mechanistic understanding of carcinogenic phenomena into their assessments (Rodricks, 2003). The saccharin story broke in 1977 and increased these attacks on the regulatory community; much public ridicule was directed at regulatory approaches, which take seriously animal carcinogenicity of the type available on saccharin, and project low-dose risks using "unverified models." Scientific debates truly raged during this time. Those who advocated risk-based approaches were accused of distorting science, of adopting scientifically suspect and untested hypotheses about dose-response relations outside the observable range and about quantita-tive measures of interspecies differences in response. Indeed they were attacked for relying upon high-dose animal experiments at all. The debates were often constructive, but it was clear that there was much unease in the scientific community and among the public about these regulatory developments.

The regulatory agencies felt, during the early days of the Carter Administration, that there might be value in developing collaborative projects on major topics of mutual concern, and to offer to the public consistent and uniform programs to deal with them. One project concerned quantitative risk assessment, and the heads of the regulatory agencies involved—Douglas Costle at the EPA, Donald Kennedy at the FDA, Eula Bingham at the OSHA, Susan King at the CPSC, and Carol Tucker Foreman at the USDA—asked that a committee of what was called the Interagency Regulatory Liaison Group (IRLG) be organized to develop uniform guidelines for the conduct of carcinogen risk assessment. The committee included scientists from the regulatory agencies (I was asked to Chair) and also leading investigators from the National Cancer Institute. We spent about a year at this effort and published our guidelines in 1979. The agencies would regulate carcinogenic risks according to their respective legislative mandates, but they would assess those risks in the same way, according to the IRLG guidelines (IRLG, 1979; Merrill, 2003).

The success of the IRLG guidelines effort did not, of course, put the scientific and policy debates to rest. Indeed, the U.S. Congress entered the fray and set in motion a National Academy of Sciences study of risk assessment in the federal government. The Academy was to form a committee to examine federal practices, and also to evaluate whether a separate and independent governmental institution should be established to conduct risk assessments on behalf of all of the regulatory agencies (Johnson and Reisa, 2003).

This question of institutional separation arose because of allegations that risk assessment outcomes could be easily manipulated to meet the predetermined desires of regulators—to regulate or not to regulate, depending upon the political climate and other factors unrelated to public health. Institutional separation of the scientific activities of regulators from those of the policy makers should help to purify those activities, or so some thought.

THE 1983 NATIONAL ACADEMY RED BOOK

"Risk Assessment in the Federal Government: Managing the Process" was published in 1983. It is a small red-covered book and it is still essential reading (NRC, 1983). Its recommendations have had a profound influence on regulatory and public health policies throughout the world. I think that those of us who served on the Committee (which was chaired by the prominent epidemiologist, Reuel Stallones of the University of Texas), and who were involved in various forums which were held soon after its release, felt almost immediately that both the regulatory community and its critics would recognize its value. And, although there is plenty of evidence that some of its lessons have been misunderstood (either willfully or more likely because of inattention), its major lessons remain influential. I shall in the later sections of this paper discuss the Red Book's influences in relation to the science of toxicology and its role in risk assessment, but, before I do that, I want to describe some of the central ideas that emerged from it, and then offer a birds-eye-view of the 25 years since its publication.

The Red Book made plain the reasons why risk assessment, even considering its inherent uncertainties, is essential to rational regulatory and public health decision making. While the report's focus was on risks from chemical toxicity, it also made plain that all types of threats to health could be evaluated within a consistent and systematic framework. Risk assessment is that framework, and its four well-known steps were first described in the Red Book. The final, integrating step, called risk characterization, is (if an assessment is scientifically adequate) a description of what is known and what is unknown about the risk—a depiction of uncertainty is an essential feature of all proper risk characterizations, just as it is in all areas of science.

Risk assessments are conceptually distinct from research; they are not sources of new knowledge, but instead they serve the important purpose of clarifying our state of knowledge at a given point in time. Because, if properly conducted, risk assessments are so highly systematic, they are almost perfect guides to research—they can elucidate what additional information and knowledge will be most important for improving our understanding, both on a chemical-specific and on a more general basis. Risk assessments are also conceptually distinct from what the Red Book described as risk management. The latter involves decisions about whether actions are needed to reduce risks in specific circumstances, the degree of risk reduction required, the means for achieving

the reduction, and the justification for all of that. Risk managers need to be in constant communication with risk assessors, so that the assessments are developed in their proper contexts, but managers do not in any way influence the scientific conduct of the assessments.

No assessment can be completed without the imposition of certain assumptions that do not have complete scientific substantiation. There is a clear danger, the Red Book authors announced, that assessments can be manipulated by the assessor (whether under the influence of the risk manager or not), to select, on a case-by-case basis, those assumptions that will guarantee a desired outcome. It is for this reason that agencies should develop technical guidelines for the conduct of risk assessments, and also specify in those guidelines the assumptions to be used to deal with gaps and uncertainties in scientific knowledge. The selection for risk assessment of a specific set of toxicological responses out of many that are known, or of animal study data in favor of epidemiology data, or of specific models for high-to-low dose and interspecies extrapolation, or of specific uncertainty factors to deal with variability in response, are all examples of the choices that need to be made to complete risk assessments and that have profound influences on their outcomes. In every case, there are several different choices that might do the job, and in the absence of highly certain scientific knowledge, an assessor, whether directed to do so by a manager or not, can make specific choices that would guarantee a predetermined and desired outcome. If agencies develop explicit guidelines that specify the choices (an assumption, or a model of some kind) that will consistently be used, then case-by-case manipulations can be avoided or at least minimized. The selection of generic assumptions and models involves a scientific evaluation of the relative merits of the alternatives available, and then a policy judgment to select the one to be used generically. Policy judgments play a role in risk assessments, but they are different in kind from those involved in risk management. Generic assumptions that are used in risk assessments have come to be called "defaults."

And, to come to one final and significant Red Book recommendation, agency guidelines should be flexible, and allow for alternative assumptions or models in specific cases in which data become available to show the alternative has stronger support than the usual default. This notion of flexibility was a central feature of the Red Book's recommendations, but it has been very difficult for regulatory risk assessors to decide when defaults can be replaced. This last issue will be examined when I reach my final sections, because toxicology is the primary engine for moving risk assessments away from defaults.

All the above emerged from the Red Book, and did not exist or existed in only partial, fragmented form, without any consensus, prior to the Red Book's publication. Moreover, the conduct of risk assessment came to be seen, under the Red Book framework, as requiring integration of several scientific disciplines—epidemiology, toxicology in all of its many manifestations, mathematical modeling, exposure assessment, and the evolving discipline of uncertainty analysis—it fostered new and valuable professional relationships and associations. This can be readily observed at any meeting of the Society for Risk Analysis (which I helped to organize in 1980). Furthermore, the ideas that have emerged from the risk assessment community are now very much in evidence at professional toxicology meetings and in the toxicology literature (Johnson and Reisa, 2003).

As a final note, I should mention that the Red Book committee firmly rejected the proposal to establish an independent institution that would serve up scientifically "pure" risk assessments to the regulators. The Committee held that the much less drastic means described in the foregoing would be adequate to prevent the improper tainting of risk assessments, and, furthermore, that close communications among the developers and users of risk assessment were essential to retain, and could well be lost were institutional separation to be enforced.

The EPA has been by far the most active agency in developing technical guidelines that generally adhere to the principles outlined in the Red Book (Anderson, 2003). Although other agency practices seem generally to follow those EPA guidelines, efforts to develop government-wide guidelines are nowhere in sight. This is regrettable; indeed, international harmonization of risk assessment guidelines would seem highly desirable.

There have been numerous National Research Council studies of risk assessment since the Red Book, and those studies generally reinforce and clarify the thinking first set forth in it (OMB 2006). Much of the latter work has emphasized some of the social and even psychological aspects of risk, and the importance for risk management of understanding that most people do not perceive threats to their health merely as matters of probabilities. Perceptions are influenced by many factors, and in most cases these are not irrational, but rather reflect common features of human psychology. In any case, the continued advance of risk assess-ment, the effort to improve its conduct and to make explicit the scientific and policy bases for certain choices in the process, and the acknowledgement of the need for incorporation of defaults, or science-based alternatives to them, are all clearly in evidence. Although the record is far from perfect, the trends seem to be good ones (Rodricks, 2001; Goldman, 2003; Rodricks, 2006).

At the same time risk assessment remains a troubled undertaking. If we focus on the troubles that are purely scientific in nature, we see that much has to do with the science of toxicology.

HOW CAN TOXICOLOGY IMPROVE RISK ASSESSMENT?

All risk assessments require that inferences be drawn from health risk information obtained under certain conditions, by appropriate scientific methods, to allow risks to be characterized under different conditions. More specifically, inferences must be drawn from effects observed in one type of population—human or experimental—to characterize effects in a second type of population—human only—that has not been subjected to direct investigation. The types of inferences needed depend upon the known differences in the characteristics of the two populations, and the known differences in the conditions under which the two populations are exposed; the larger these differences, the greater the number of inferences needed to complete a risk assessment, and the less certain the final risk characterization. The Red Book Committee used the term "inference options" to describe the several that might have applicability for a specific type of necessary inference; the one selected for general use, in the absence of specific data to support a scien-tifically more meritorious one, I have referred to as a "default." The greater the difference between the conditions under which health risk information is obtained by actual study and the conditions of interest for risk assessment, the greater the numbers and types of inference options to be considered and defaults used.

Two simple examples illustrate this last point. In the first example, we obtain relatively clear informa-tion from a series of occupational epidemiology studies that reveal a consistent dose-risk pattern relating exposures to benzene to excess rates of acute leukemia. Our occupational health officials can apply that dose-risk information to other, unstudied cohorts of workers exposed to benzene at similar levels with little need for the types of inferences I have mentioned. It will probably not be a completely "default-free" risk assessment, however, because there may be small differences in the magnitudes of exposure in the studied and unstudied populations, and in the age and gender distributions in the two populations (a difference that we generally do not know how to accommodate, thus introducing the need for a default assumption).

Now consider a case in which a compound that has not been subjected to significant epidemiological study has been shown to produce excess tumors in one but not a second animal species. Our population of interest is not occupational, but includes every type of individual found in the general population. It goes without saying that the numbers and types of inferences needed to reach any conclusions about health risks in such a population are both far greater than what is needed in the first case. What I have said is not, of course, limited to carcinogenic responses, but applies to any type of toxicity. In cases such as this second one, and even in many less extreme, defaults may play a greater role in risk assessments than do scientific knowledge and data. The advantages in such default-driven risk assessments relate to the high degree of

consistency their use brings as we move from one risk assessment to the next, and to their value in reducing the chance of case-by-case manipulations of assessments to yield risk results the manager would prefer to have. The clear disadvantage of default-driven risk assessments is the lack of a basis for any claim that they provide reliable knowledge about risks.

Many argue, of course, that we really do not need reliable knowledge about risks to health, except for those risks that are very large, affect large numbers of people, and which can be directly measured with the tools of epidemiology and clinical science. For all other sources of risk we need only a system for identifying hazards (the animal bioassay, the standard epidemiology study), and a set of practical tools (the aforementioned "defaults") for specifying levels of human exposure that are likely to be without significant risk. If we can be reasonably confident that those tools do not lead us to underestimate human risk, then the search for reliable knowledge is an unnecessary luxury (Goldman, 2003; Lupien, 2002).

Under this last view, testing to identify the toxic properties of chemicals becomes the principal goal of toxicology, toward the end of regulating increasingly large numbers of chemicals as quickly as possible, with minimum knowledge of how they actually affect health. I am sure most toxicologists do not see their science as having such narrow bounds.

Indeed, toxicologists seem intent on developing and applying methods that allow statements about health risks to be made with increasingly greater reliability. If they are to contribute significantly to improved reliability, they should seek to develop the combinations of experimental models that can yield information about human hazards with demonstrably greater predictive power, and with greater efficiency. They should seek to understand the relationships between so-called "toxicity end points" and actual human diseases. They should seek to understand the behaviors of dose-response relationships in the region of human exposures. They should seek to understand all of the factors that contribute to variability in response among individuals, and to quantify their cumulative effects, and they should seek to develop reliable profiles of the distributions of risk among individuals and to learn whether there are subpopulations of high susceptibility individuals that are described by a different distribution. They should understand that close cooperation with epidemiologists and molecular biologists and biologically oriented mathematicians is necessary for all this. They should hope that the results of all this research could lead to some general rules (*i.e.*, models) about these and other critical aspects of risk assessment; in other words, they hope that toxicological risk assessments can become as reliable as the physical and chemical sciences.

What I have just described, if it could be realized, would immensely improve our knowledge of how chemicals in our environment and in the thousands of products with which we come into contact, both of natural and industrial origin, contribute to human disease. In other words, this type of toxicological knowledge is necessary if the goals of developing scientifically reliable risk assessments are to be achieved in full.

These ambitions can be achieved only very slowly, with the gradual accumulation of small increments in knowledge, and then with periodic syntheses, directed toward developing certain general (though always tentative) truths. Because toxicology has become such a dynamic and diverse set of scientific activities, it is difficult to make out its general directions, and to know whether the goals that need to be achieved to allow the development of scientifically reliable risk assessments are likely to emerge at a sufficiently rapid pace to allow this model of toxicology to compete with the model that emphasizes testing over reliability. If we are smart, we will want to identify ways to pursue both models simultaneously!

It will take individuals with broad understanding and experience, and strong creativity, to recognize when we are ready for new syntheses of knowledge and to reveal to us the form and utility of that knowledge. These types of syntheses of the state-of-scientific knowledge typically emerge from studies of the National Academy of Sciences and similar institutions that have the capacity to bring together committees of leading experts (OMB, 2006). If one traces the evolution of such expert committees over the past 25 years, it is seen that they are increasingly multidisciplinary in nature, and recognize that coordinated efforts among all the many scientists and medical experts who study environmental determinants of human health are necessary to advance understanding.

If one studies the many reports these expert committees have written, it is possible to discern a strong and evolving belief that the path to reliability in risk assessment entails a deeper and broader look at toxicological mechanisms. The central unresolved questions in risk assessment, which I have outlined above, will, it is widely assumed, eventually yield to the pursuit of greater and greater understanding of the pharmacokinetic and pharmacodynamic phenomena underlying the induction of adverse health effects. Such improved understanding creates the opportunity for improved reliability in risk assessments. If we plan and follow this path, and if we are lucky, we should simultaneously be able to achieve the desirable goal of more rapid (and perhaps even whole-animal-free) testing to identify the hazardous properties of increasing numbers of chemicals. Development and validation of in vitro assays, some involving the remarkable tools of toxicogenomics, would seem to be the keys to understanding toxicity pathways, and to using that understanding to develop the pharmacokinetic and pharmacodynamic models necessary to predict human risk with greater reliability. At the same time toxicologists need to link this vision to developments in epidemiology and certain types of population studies, most especially biomonitoring efforts, because these types of studies provide the only opportunities for testing our experimentally generated hypotheses about human risks.

The continuing syntheses of knowledge in the science of toxicology is increasingly undertaken within the risk assessment framework; that is, our leading experts seem consistently to be asking how the methods and data emerging from the toxicology laboratory create opportunities to make more efficient and more reliable our quest to understand the health risks—indeed, the human diseases—associated with inadequately controlled exposures to chemical products and contaminants. This is to be encouraged, because it is the first and necessary step toward the actions that are necessary to genuinely improve the public health.

HOW IS TOXICOLOGY IMPROVED BY RISK ASSESSMENT?

For all of us engaged in the daily struggles to understand toxic phenomena and human health risks, we can continue to learn much from the thinking and goals the risk assessment community has set before us over the past three decades. We have a clearly organized scientific framework within which we can evaluate both what we understand and do not understand about the health consequences of chemical exposures. Adequate descriptions of uncertainties are demanded from all of the disciplines, including toxicology, that contribute to risk assessments. Uncertainty analysis remains an immature science; it is easy simply to list all that we do not understand, but it is not at all easy to provide the type of systematic and even quantitative descriptions of uncertainties that can be useful to decision-makers, and that, at the same time, avoids confounding decision-making by presenting unnecessarily complex analyses (Finkel, 2003). We need to do much to improve our handling of uncertainties in risk assessment, and we also need to work closely with risk managers and risk communicators to ensure that understanding is sufficient on all sides to support good decision-making. A large part of the burden of describing and quantifying uncertainties falls upon toxicologists, and we should be looking to the larger risk assessment community for guidance on this topic.

This type of communication between assessors and managers of risk has always been strongly advocated by the risk assessment community. The need for it extends beyond the question of scientific uncertainty and its role in decision-making, and includes the need for toxicologists and other scientists involved in risk assessment to understand fully the regulatory or public health contexts in which risk questions arise (NRC, 1994). If toxicology research is to have the effects we would all like it to have, then toxicologists need to understand the needs of risk assessors and managers, and to be ready to respond to the significant uncertainties in risk assessments, both general and chemical-specific.

The risk assessment process is very demanding, and requires rigorous and quantitative scientific methods. If defaults are to be avoided, then toxicologists must be able to provide alternative approaches that can claim a degree of certainty that is consistent with the current state of the science. Here, by the way, toxicolo-

gists should demand of risk assessors, and the managers they serve, less rigidity of thinking than we have historically witnessed. That is, we should hope that rigorous new science could replace defaults even if we cannot claim it provides absolute certainty (which, of course, no scientific finding can). I am not accusing risk assessors in regulatory agencies of seeking absolute certainty, but it seems that the path to replacing defaults is poorly defined and often seems nearly impassable. I would argue that the judgment about whether new scientific information is "sufficiently certain" to replace an existing default is only partly a scientific one: in fact, I would further argue, the scientific task in such situations is the very difficult one of describing the degree of reliability of new information, but does not extend to the policy arena that judges whether that degree of reliability is sufficient to replace a default (Brown and Rodricks, 1991). We need a new model for decision-making in this context. In any event, toxicologists need to be attuned to these types of problems, and become much more proficient at describing the certainties and uncertainties in any findings that they think should influence the risk assessment process.

Risk assessors think probabilistically. This type of thinking is difficult for most people, including many scientists. The model for decision-making that has evolved for carcinogens that have not been shown to display thresholds, involves such probabilistic thinking. Using probabilistic analyses in decision-making is, I would hold, highly desirable, because their use provides insight into how various actions actually affect risks, and they allow decision-makers some flexibility in risk decisions—there is no "fixed" definition of "safe" or "acceptable" exposures under this decision model.

The models we have for non-cancer endpoints, and for carcinogens that appear to display thresholds, are deficient in this respect, and I suggest we begin thinking about probabilistic expressions of their health risks. The approach would involve attempts to describe the distribution of thresholds in a population, together with estimates of the distribution of human exposures. The degree to which the two distributions overlap can be used to assess the probability that thresholds will be exceeded. This is not new science; the risk literature contains several excellent examples of how this might be accomplished (Hattis and Goble, 2003). But moving more fully to probabilistic approaches will no doubt encounter resistance from those who are satisfied with the current deterministic ("yes/no") approaches. Those who are satisfied perhaps do not appreciate the difficulties encountered by decision-makers who have come to believe that the "bright line" approach to public health protection that is represented by the use of RfDs or TDIs is their only decision-making option. I believe that probabilistic risk assessment for all toxicity end points, coupled with careful descriptions of their attendant uncertainties, is the appropriate future for our science. I should add that I do not propose the probabilistic approach to replace the current system of safety assessment for deliberately introduced substances—pesticides, food ingredients, substances used in consumer products—that require pre-market evaluation, but I do think it necessary for the more flexible approach to decision-making that is needed for environmental contaminants of every type.

Toxicology is an extremely rich and challenging science and is also of enormous importance to the public's health and to the economy. We need to continue to improve all the processes whereby decisions are made regarding the controls we place on exposures to chemicals, and the risk framework that has evolved over the past three decades should remain a dominant feature of those processes. We might think of toxicology as having no more important function than the improvement of risk assessments, and we might also see no better guide to the advance of toxicology than what is revealed by the conduct of risk assessments within the rigorous framework that has, I hope, now become a firmly entrenched feature of the thinking of all who are concerned about science-based public health protection.

My journey from those early days spent among the mycotoxins—indeed from the even earlier days when I journeyed from a small Massachusetts town to take up study at MIT (an experience both thrilling and terrifying)—my professional life has been one long and enormously gratifying experience in learning. My 15 years at the FDA provided me completely unexpected opportunities for establishing credentials in science and in regulatory policy, and I could not imagine improving on the twenty-five years I have worked as a consultant at ENVIRON—where I have had such satisfying personal and professional relationships with a seemingly endless number of stimulating people both within the company and among its many clients. I can not begin to name all those many individuals whose professional lives have intersected with mine, and who have, knowingly or unknowingly, contributed to my understanding of the science to which we are all devoted, and to that difficult, and sometimes ill-defined boundary between science and the world of policy. The social, economic, and political importance of the way we develop and interpret toxicological information makes controversy—indeed conflict—inevitable, and this dimension of our discipline becomes most manifest when we make the transition from the research setting into the domains of risk assessment, management, and communication. If I have made any contribution to toxicology, I hope it has been to help ensure that the transition remains faithful to what the underlying science has revealed (and failed to reveal), and yet yields risk assessments that have practical utility and public acceptance. My achievements in this area fall short of their goals, but I have immensely enjoyed the challenge; I am not ready to retire from it.

I have been both gratified and somewhat surprised by this award from the ACT. It puts me in the company of a truly distinguished group of toxicologists, and I could wish for no better company. I am somewhat surprised because I have not made significant contributions to the life of ACT—indeed to that of any professional society. My only excuse for this lack of contribution stems from the fact that most of my donated professional time since the mid-1970's—when I became a member of the National Research Council's Committee on Toxicology, then under the intellectually sparkling tutelage of Joseph Borzelleca, a person who has been a mentor to so many of us—has been taken up by such NRC work. In a few weeks I will begin service on my 24th NRC or IOM Committee, this one devoted to a new look at EPA risk assessment practices. Even while finding all this committee work immensely satisfying, I have always felt a little guilty about not having the time to assist with the many important ACT activities. Perhaps this award—for which I can find no better words than "thank you"—will inspire me to find that time in what I hope is a long future in "the science of poisons."

REFERENCES

Albert, R.E., Train, R., and Anderson E. 1977. Rationale developed by EPA for the assessment of carcinogenic risks.
J. Nat. Cancer Ins. 58(5): 1537-1541.

Anderson, E.L. 2003. The Red Book in context: science at the center. *Human Ecol. Risk Assess.* 9(5):1197-1202.

Armitage, P. and Doll, R. 1954. The age distribution of cancer and a multistage theory of carcinogeneses.
Br. J. Cancer. 8:1-12.

Brown, S.L., and J.V. Rodricks. 1991. A graphical display of risk information. In *Risk Analysis,* ed. C. Zervos. 45-53.

Crump, K. 1976. Fundamental carcinogenic processes and their implications for low-dose risk.
Cancer Res. 36:2973-2979.

EPA (U.S. Environmental Protection Agency). 1986. Guidelines for carcinogen risk assessment.
Fed Reg. 51(185):33992-34008.

FDA. 1977. Chemical compounds in food-producing animals; criteria and procedures for evaluating assays for carcinogenic residues. *Fed. Reg.* 42:10412-10429.

Finkel, Adam M. 2003. Too much of the "Red Book" is still (!) ahead of its time. *Human Ecol. Risk Assess.* 9(5):1253-1271.

Fitzhugh, O.G. 1949. Procedures for the appraisal of the toxicity of chemicals in foods, subacute and chronic toxicity.
Food Drug and Cosmetic Law Quarterly. 421-425.

Goldman, Lynn R. 2003. The Red Book: A reassessment of risk assessment.
Human and Ecological Risk Assessment. 9(5), 1273-1281.

Hattis, D. and Goble, R. 2003. The Red Book, risk assessment, and policy analysis: the road not taken. *Human Ecol. Risk Assess.* 9(5):1297-1306.

Hutt, P.B. and Merrill, R. 1991. Regulation of Carcinogens. Chapter VII. in *Food and Drug Law.* 2nd Ed. The Foundation Press, Inc, Westbury, NY. 863-963.

IRLG (Interagency Regulatory Liaison Group). 1979. Scientific bases for identification of potential carcinogens and estimation of risks. *J. Nat. Cancer Inst.* 63:241-268.

Johnson, Barry L. and James J. Reisa. 2003. Essays in Commemoration of the 20th Anniversary of the National Research Council's *Risk Assessment in the Federal Government: Managing the Process. Human Ecol. Risk Assess.* 9(5):1093-1099.

Lehman, A.J., Laug E.P., and Woodard G. 1955. Procedures for the appraisal of the toxicity of chemicals in foods, drugs, and cosmetics. *Food Drug Cosmetic Law Journal.* 10:679-748.

Lupien, J.R. 2002. The precautionary principle and other non-tariff barriers to free and fair international food trade. *Crit Rev Food Sci Nutr.* 42:403-415.

Mantel, N. and Bryan, W. 1961. "Safety" testing of carcinogenic agents. *J. Natl. Cancer Inst.* 27:455-470.

Merrill, Richard A. 2003. The Red Book in historical context. *Human and Ecol. Risk Assess.* 9(5):1119-1127.

NRC (National Research Council). 1983. *Risk Assessment in the Federal Government: Managing the Process.* National Academy Press, Washington, DC.

NRC (National Research Council). 1994. *Science and Judgment in Risk Assessment.* National Academy Press, Washington, DC.

OMB (Office of Management and Budget). 2006. *Proposed Risk Assessment Bulletin.* The proposed bulletin is an excellent source for identifying the many NRC studies pertaining to risk assessment.

Rodricks, J.V., Hesseltine, C. and Mehlman, M. (Eds.) 1977. *Mycotoxins in Human and Animal Health.* Pathox Pub. Chicago.

Rodricks, J.V. 1981. Regulation of carcinogens in food. *Annals of the New York Acad. Sci.* 363:29-35.

Rodricks, J.V. 1990. Risk assessment and animal drug residues. *Drug Metabolism Reviews.* 22(6-8):601-607.

Rodricks, J.V. 2001. Some attributes of risk influencing decision-making by public health and regulatory officials. *American Journal Epi.* 154(S) (12):S7-S12.

Rodricks, J.V. 2003. What happened to the Red Book's second most important recommendation? *Human Ecol. Risk Assess.* 9:1169-1180.

Rodricks, J.V. 2006. *Calculated Risks: The toxicity and human health risks of chemicals in our environment.* 2nd Ed., Revised. Cambridge University Press. Cambridge, England.

Stoloff, L. and J.V. Rodricks. 1977. Aflatoxin residues from contaminated feed in edible tissues of food-producing animals. In *Mycotoxins in Human and Animal Health*, Rodricks, J.V., Hesseltine, C. and Mehlman, M. (Eds.) Pathotox Pub. Chicago. pp 67-80.

Zervos, C. and J.V. Rodricks. 1982. FDA's ban of DES in meat production. *The American Statistician* 36:278-283.